New insights into the evolutionary history of geminiviruses derived through the discovery of divergent viruses isolated from wild plants

P. Bernardo,[1,2] M. Golden,[3] A. Mohd,[4] E. Fernandez,[1] M. Granier,[1] A.G. Rebelo,[5] M. Peterschmitt,[1] D.P. Martin[3] and P. Roumagnac[1]

[1] CIRAD/UMR BGPI TA A54/K Campus International de Baillarguet 34398 Montpellier Cedex 5, France (e-mail: philippe.roumagnac@cirad.fr)
[2] INRA/UMR BGPI TA A54/K Campus International de Baillarguet 34398 Montpellier Cedex 5, France
[3] Computational Biology Group. Institute of Infectious Disease and Molecular Medicine. UCT Faculty Of Health Sciences. Observatory 7925. South Africa
[4] Division of Crop Protection, Indian Institute of Pulses Research, Kalyanpur, Kanpur-208024, India
[5] South African National Biodiversity Institute, Kirstenbosch, Private Bag X7, Claremont 7735, Cape Town, South Africa

During a large scale "*non a priori*" survey in 2010 of South African plant-infecting single stranded DNA viruses, a highly divergent geminivirus genome was isolated from an uncultivated spurge, *Euphorbia caput-medusae*. In addition to being infectious in *E. caput-medusae*, the cloned viral genome was also infectious in the cultivated hosts, tomato and *Nicotiana benthamiana*. The virus, named Euphorbia caput-medusae Latent virus (EcmLV) due to the absence of infection symptoms displayed by its natural host, caused severe symptoms in both of the cultivated plant species.

The genome organization of EcmLV is unique amongst geminiviruses and it likely expresses at least two proteins without any detectable homologues within public sequence databases. Although clearly a geminivirus, EcmLV is so divergent that we propose its placement within a new genus that we have tentatively named Capulavirus.

Using the most divergent set of geminivirus genomes ever assembled, we detect strong evidence that recombination has likely been a primary process in the genus-level diversification of geminiviruses. We demonstrate how this insight, taken together with phylogenetic analyses of predicted coat protein and replication associated protein (Rep) amino acid sequences indicate that the most recent common ancestor of the geminiviruses was likely a dicot-infecting virus that, like modern day mastreviruses and becurtoviruses, expressed its Rep from a spliced complementary strand transcript.